AD-A058 478 DECISIONS AND DESIGNS INC MCLEAN VA F/G 5/1

ADVANCED DECISION TECHNOLOGY PROGRAM.(U)

DEC 77 N00014-76-C-0074

UNCLASSIFIED PR-77-12-30 NL

1 OF 1

AD A058 478

END

DATE FILMED

11-78

DDC

(4)

1 October 1976 – 30 September 1977

ADVANCED DECISION TECHNOLOGY PROGRAM

ANNUAL PROGRESS REPORT PR 77-12-30

DECISIONS AND DESIGNS INCORPORATED

ADVANCED ARPA DECISION TECHNOLOGY PROGRAM

CYBERNETICS TECHNOLOGY OFFICE
DEFENSE ADVANCED RESEARCH PROJECTS AGENCY
Office of Naval Research • Engineering Psychology Programs

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: PR 77-12-30	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): ADVANCED DECISION TECHNOLOGY PROGRAM ANNUAL PROGRESS REPORT	5. TYPE OF REPORT & PERIOD COVERED: Periodic; 1 Oct 1976-30 Sept. 1977
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s)	8. CONTRACT OR GRANT NUMBER(s): N00014-76-C-0074
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Decisions and Designs, Incorporated, Suite 600, 8400 Westpark Dr., P.O. Box 907, McLean, VA 22101	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS: Defense Advanced Research Projects Agency, 1400 Wilson Boulevard, Arlington, VA 22209	12. REPORT DATE: December 1977
	13. NUMBER OF PAGES: 58
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office): Office of Naval Research, Code 455, 800 North Quincy Street, Arlington, VA 22217	15. SECURITY CLASS. (of this report): UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

decision theory research
multi-attribute utility assessment
decision making
decision aiding
human judgment
judgmental biases

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report presents a review of salient progress and accomplishments achieved in the ARPA-supported Advanced Decision Technology Program during Fiscal Year 1977. The research highlighted is concerned with a variety of technical questions fundamental to the development and refinement of utilitarian decision aids. Applied research activities conducted under the research program concerned with the application of the variety of decision aiding concepts to DOD decision problems are also reviewed.

4

1 OCTOBER 1976 — 30 SEPTEMBER 1977

6

ADVANCED DECISION TECHNOLOGY PROGRAM.

9 **ANNUAL PROGRESS REPORT. PR 77-12-30**

1 Oct 76 - 30 Sep 77.

This research was supported by the
Advanced Research Projects Agency of the
Department of Defense
and was monitored by the
Office of Naval Research
under Contract N00014-76-C-0074

15

11 30 Dec[redacted]977

12 60p.

14 PR-77-12-30

DECISIONS and DESIGNS, INC.

Suite 600, 8400 Westpark Drive
McLean, Virginia 22101
(703) 821-2828

PREFACE

This report covers highlights of technical progress achieved in the ARPA-supported Advanced Decision Technology Program over the past 12 months (1 October 1976 through 30 September 1977). The work summarized herein represents the collective efforts of Decisions and Designs, Incorporated, Perceptronics, Harvard University, Stanford University, and the University of Southern California. Much of the material presented in this report was extracted from technical information prepared by each of the program participants, principals being: C. W. Kelly, Decisions and Designs, Incorporated; Howard Raiffa and Richard Meyer, Harvard; Ronald Howard, Stanford; Ward Edwards, University of Southern California; and Paul Slovic, Sarah Lichtenstein, and Baruch Fischhoff of Perceptronics, Incorporated.

SUMMARY

This report presents a review of salient progress and accomplishments achieved in the ARPA-supported Advanced Decision Technology Program during Fiscal Year 1977.

In the basic research elements of the program, experimental work moved ahead addressing a wide variety of technical questions fundamental to an understanding of decision behavior and to the development and refinement of utilitarian decision aids. In research conducted during the past 12 months, the fallibility of unaided human judgment in decision contexts was further underscored through the discovery of additional judgmental biases, and research on methods to ameliorate or eliminate these biases was advanced. Valuable new insights were gained about how best to aggregate group judgments, about which attribute weighting methods are most appropriate in multi-attribute utility assessment contexts, and how accurately people assess the true causes of events. Research was also advanced concerning new approaches to the estimation of very small probabilities, the validation and simplification of multi-attribute measurement methods, and optimal methods for eliciting quantified subjective judgments.

In the applied portion of the program, pilot application efforts met with considerable success in achieving goals of user familiarization, technology transfer, and decision aid evaluation. The pilot application work at Headquarters, EUCOM was extended to the subordinate commands (NAVEUR, USAFE, USAREUR) with continued favorable user interest. Multi-attribute utility assessment methods were successfully applied over the past 12 months to a number of Army system evaluation problems, to the problem of assessing the combat readiness of Marine Corps Units, and to resource

allocation choices in the formal USMC budgeting process (POM). A number of analytical efforts were also completed during the past contract period concerned with decision aid evaluation, methodology for the assessment of command and control systems, the use and abuse of formal analytic methods in public policy decisions, and critical assessment of the state-of-the-art of decision analytic methodology and its application.

CONTENTS

1.0 INTRODUCTION

The Advanced Decision Technology Program has as its objective the development of improved methods for enhancing the timeliness, accuracy and efficiency of Defense decision functions and the transfer of these technologies to active use in DoD. To those who have seriously grappled with important problems of choice in their personal spheres (such as purchasing a home, deciding which automobile to buy, choosing among job options, etc.) the difficulties inherent in decision making are apparent. Even at the relatively uncomplicated level of personal choice, we often encounter more relevant decision dimensions than the intellect can cope with; many dimensions are difficult to value in an objective way, and a highly uncertain world is often interposed between possible choices and their outcomes.

While most of us cope with personal decisions with more or less systematic consideration (and usually with unknown degrees of success or failure), it takes little extrapolation to realize that in national security decision contexts, the problems are often far more complex, the uncertainties greater, and the stakes involved of enormous magnitude. These factors, coupled with a research base that points to suboptimal human performance in complex decision tasks, serve as the prime motivation for this program of research. There is a serious need to understand better how humans reach decisions, what their strengths and weaknesses are in decision contexts and to develop from that knowledge base techniques to aid Defense decision makers in reaching better choices.

In keeping with the above, the Advanced Decision Technology Program encompasses a range of research activities from basic research on factors influencing human judgments

to applied activities concerned with the development and trial application of decision aiding methods. For the most part, the formal methodology of decision analysis is a central theme of the research project. This normative decision-aiding model, and variants of it, are the basis for applied activities conducted within the program, and the behavioral requisites of this model provide a general rubric for the related basic research activities.

The Advanced Decision Technology Program achieved highly significant progress during the FY 1977 contract period. Participating scientists at Decisions and Designs, Incorporated, Harvard, Stanford, the University of Southern California, and Perceptronics, Incorporated, have made noteworthy contributions to our knowledge of human decision processes, to the methodological base for decision-aiding systems, and to the development and application of decision aids in Defense decision contexts. In the following sections, highlights of FY 1977 research activities supported under the program are presented.

2.0 SUMMARY OF PROGRESS AND ACCOMPLISHMENTS

2.1 FY 1977 Progress: Research on Decision Processes and Aiding Technologies

This program area subsumes a wide variety of basic research tasks focused on improving our understanding of human decision-making performance, on determining the factors that serve to enhance or degrade decisions, and on advancing the methodological substrate for decision-aiding technology. Representative accomplishments and progress in this program area are presented in the following paragraphs.

2.1.1 Evaluation of past decisions - Learning derived from prior experience is a key element in decision making. Whether in the context of casual, intuitive decision making or in meeting the requirements of formal decision models, the veridicality of judgment brought to bear is largely a product of years of experience. This reality has led researchers concerned with decision behavior to inquire into the factors that influence learning from experience.

Research initiated in FY 1976 within this general theme revealed serious deficiencies in the way people evaluate the wisdom of their past decisions, and those of others. The research demonstrated a strong hindsight bias, i.e., a strong tendency on the part of people to exaggerate the predictability of past events. Given knowledge about how events actually turned out, people find it virtually impossible to assume the same uncertainty set that prevailed at the time the prior decision was taken. In addition to seriously distorting judgment of the acumen of past decisions (often with unfortunate consequences to the decision makers), the bias clearly limits what people learn from prior actions and consequently limits also the perspective that can be brought to bear on current choices.

In extensions of this avenue of research during FY 1977, Shaklee and Fischhoff [reference 30] conducted a series of experiments concerned with how people decide on the adequacy of explanation of events that have occurred. These experiments were designed to compare rational models of multi-causal explanation with a postulated principle of nonrational minimum causation. The latter principle states that once an event is sufficiently explained, other possible causes are seen as less likely to be involved. The experimental results clearly support the principle of minimum causation. Explanations for an event are apparently sought until the event is plausibly explained. Once the event is thus explained, other possible and relevant causes are considered superfluous.

This behavior pattern can certainly distort the level of understanding derived from experience. Since the order in which information is received about an event is often happenstance, chance may play a major role in determining which possible cause is given main causal weight and which causes are discounted. This phenomenon could have important consequences to information seeking. Knowing that one cause was definitely involved may tend to make people uninterested in searching for additional causal evidence. This could be particularly damaging when the first causal evidence is either unreliable or erroneous. In such cases, true causality may never be known because the question was prematurely closed. These results, along with those from earlier work in this area, serve as the basis for a major theoretical paper on the psychology of explanation which is now in preparation.

2.1.2 Structuring decision problems - An important class of decision problems entails system diagnosis: troubleshooting of existing systems and mode of failure analysis of projected systems. These types of decision problems are

commonly represented by fault trees which depict interdependencies within the system and possible sources of failure. Once constructed, the tree serves as a guide to the problem solver in assessing the probability of given types of failure, in identifying faults in failed equipment, and in pinpointing probable causes of failure.

As with many other kinds of problems, fault trees can be presented in a number of different ways. In FY 1977, research was completed which looked at the impact of three such discretionary aspects of the way in which fault trees are structured on the resulting evaluations produced [reference 14]. The structural differences taken under study included: a) what is specifically delineated in the fault tree and what is categorized as "all other," b) the amount of detail presented in the various branches of the fault tree, and c) how various subsystems are grouped into branches. The principal results of these studies showed that: a) people are quite insensitive to what has been left out of a fault tree; b) increasing the amount of detail for the tree as a whole or just for some of its branches has relatively little effect on perceptions; and c) the importance of a particular branch can be increased by presenting it in pieces (i.e., as two separate component branches). Insensitivity to omissions was found both with naive subjects and experts (in this case experienced mechanics). Such results have important implications for: a) how to best inform decision makers about risky situations and b) how experts should perform fault tree analyses.

2.1.3 Biasing factors in probabilistic judgments - Effective decision making in the face of uncertainty, whether by unaided intuitive means or in the context of formal decision-aiding models, requires that the decision maker assess relevant probabilities and values to a reasonable level of accuracy. Although it has been demonstrated that

some decision problems are relatively insensitive to substantial variations in probabilities and values, it remains the case that ill-considered, erroneous assessments can lead to bad decisions. In recognition of this, research has been focused on how people estimate probabilities, on how accurate these estimates are, and on what factors lead to biases in probabilistic judgments. This avenue of research, now entering its fourth year under the aegis of the ARPA-supported Advanced Decision Technology Program, has identified a number of biasing factors that degrade probabilistic judgments and has generally highlighted major human frailties in this domain.

One of the most important sources of bias in unaided judgments of probability and frequency identified to date is the "availability bias." Availability bias arises from the use of a cognitive strategy or "heuristic" whereby an event is judged likely or frequent if it is easy to imagine or recall instances of that event. In reality, instances of frequent events are typically easier to recall than instances of less frequent events, and likely occurrences are easier to imagine than unlikely ones. Thus, availability is often an appropriate cue for judging frequency and probability. However, since availability is also affected by numerous factors unrelated to likelihood, reliance on it may lead to overestimation of probabilities for recent, vivid, emotionally salient, or otherwise memorable or imaginable events.

In extensions of the work on probability estimating, Beyth-Marom and Fischhoff [reference 3] studied availability effects in a number of different contexts where availability could be (and was) directly measured. Although some of their results supported the availability hypothesis, it was found to be an oversimplification of the process of estimating category size. With one task availability was a

better predictor of subjects' estimates of category size than was the actual category size. In that situation, exhaustively listing instances from memory led to improved estimation. For a second task, however, availability did not predict subjective estimates, nor did exhaustive listing improve (or degrade) performance.

Thus, although availability obviously is an important factor in category size estimation, more research is needed on other factors that supplement or supplant availability in some situations. Some speculation about these factors is advanced in the paper. In addition, although working hard to exhaust one's memory for instances before making an estimate appears to be a simple and promising debiasing procedure, we need to know when it works. Using it where it is ineffective may increase confidence without improving performance.

The availability bias was further demonstrated in work by Lichtenstein, Slovic, and Fischhoff [reference 22] in which subjects were required to estimate the frequency of lethal events. The results of this work, consistent with the availability hypothesis, showed that people tend to overestimate the frequency of lethal events which get heavy media coverage (e.g., cancer, floods, tornadoes, fire) while underestimating the frequency of the less publicized ones. A follow-on study showed that these biases in perception could be predicted moderately well from simple indices of the amount of press coverage each cause of death receives, and the disproportionate share of coverage given to some causes. Efforts to improve the accuracy of subjects' judgments by informing them of the general nature of the availability phenomenon were unsuccessful. These results suggest that one should be extremely cautious in relying on people's estimates of the probability of less familiar unlikely events.

Additional work on probability assessment conducted during FY 1977 was concerned with the base-rate fallacy. This biasing factor occurs when people fail to consider relevant background information in estimating the probability of occurrence of a specific event. There are many situations wherein two kinds of information are relevant to developing a probabilistic inference: background or base-rate information about how things usually are in the situation of interest, and specific diagnostic or indicator information about the particular situation. Unless the diagnostic information is extremely good, the base-rate information is an important guide in assessing the likelihood of the specific event of interest. Bayes' rule is the appropriate statistical approach to combining base-rate and diagnostic information.

Bar-Hillel [reference 2] studied the base-rate fallacy and examined a number of explanations for it. The research showed that the effect is not an artifact of how responses are elicited nor of the order in which information is presented. Nor is it due to simple misreading of the problem. It cannot be attributed to inherent inability to integrate multiple sources of uncertainty. Base rates are apparently ignored because subjects feel they <u>should</u> be ignored. In essence, base rates often seem irrelevant when they should be given great weight. Thus, subjective and objective relevance have very different determinants. This research suggests some problem characteristics that seem to affect the perceived relevance of base-rate information and the likelihood that it will be ignored. One hypothesis, tested and confirmed in this study, is that base rates <u>will</u> be used if they can be readily interpreted as relating <u>causally</u> to the target judgment.

In sum, this study indicates the conditions most likely to produce the base-rate fallacy. The knowledge

obtained here, leading toward an understanding of when base rates are and are not viewed as relevant, has direct implications for training people to overcome this bias.

Another type of problem involving the combination of information arises when one must assess the probability of the conjunction of two events. When one asks, for example, "How likely am I to find a soldier in the ranks who can both do advanced computer programming and receive a top security clearance?" or "How likely is it that I will be a colonel and have four children by age 40?", one is dealing with compound probabilities (conjunctions). Earlier work has shown that people tend to overestimate the probability of such conjunctions, but has not shed light on why people are prone to such bias.

In six experiments conducted during the reporting period, the assessment of probabilities for conjunctive events in a variety of tasks was examined. Perhaps their most dramatic result was documentation of the overestimation bias in each case, with unique and frequentistic events, with dependent and independent events, with different response modes, and with substantively different kinds of events. In many cases, the bias was so large that the probability of the conjunction of two events was greater than the probability of one of the constituent events, a clear violation of the rules of inference.

Careful examination of the results revealed, however, that although the general shape of the bias was similar with each of these tasks, the cognitive mechanisms upon which subjects relied were different with different tasks. These differences were capitalized upon to suggest a set of debiasing procedures particularly suited to different tasks. A report on this research is in preparation.

The methods used to elicit probability distribution estimates are yet another source of error in probability estimates. Work by Alpert and Raiffa and Seaver, von Winterfeldt and Edwards along with that of others has demonstrated the phenomenon.

One problem with studies of probability assessment is that the available dependent measures are not completely satisfying ones. Proper scoring rules do not provide a sensitive measure of how closely an elicited probability distribution reflects the assessor's beliefs about the chances of occurrence of the events over which the distribution is assessed. Typical measures of calibration involve the proportion of events which occur historically, across distributions, that lie on fixed intervals of the assessed distribution, e.g., the proportion of events which obtain that were assessed as lying within the interquartile interval ($.25 < p < .75$). Such measures require assessments over huge numbers of distributions to be reliable. Furthermore, many interesting questions regarding the usefulness of elicitation techniques simply cannot be answered without a measure which takes into account a larger number of the important features which distinguish one probability distribution function from another.

John [reference 17] investigated the possibility of presenting subjects with a sample distribution of a random variable and eliciting the population distribution (density) from which the sample was presumably drawn. Stimuli were sticks of standard length painted blue and yellow. The length of yellow on each stick constituted the random variable. Subjects were shown three sample distributions (uniform, modal, and bimodal) of 26 sticks each.

Each subject used one of three probability elicitation procedures to convey his (her) knowledge of the

population distribution from which the sample was presumably drawn. In the fractile procedure, subjects were asked "to give a length of yellow such that a stick chosen randomly from the population just sampled will have a length of yellow less than or equal to the length you give with probability = (.99, .75, .50, .25, .01)." In the probability procedure, subjects were asked "to judge what the probability is that a stick chosen randomly from the population just sampled has a length of yellow less than or equal to (.65", 1.95", 3.25", 4.55", 5.85")." A third procedure (graph) required subjects to draw a curve, of which "the height at each point represents the relative probability that a stick drawn at random from the population will have that length of yellow."

The goodness of fit between the elicited and sample distributions was found to be a nonadditive function of assessment technique and sample distribution shape. Although the fractile procedure performed substantially worse for all three sample distributions, the relative performance of the probability and graph methods varied as a function of sample distribution. The finding that biases in probability assessment result from an interaction between the method of assessment and the shape of the distribution is an important one; the development of an experimental paradigm to adequately evaluate probability assessments is a topic worthy of further attention.

From an applied point of view, John's experiment once more suggests that the custom of using fractile techniques for assessing continuous distributions rather than any of their equally simple or simpler alternatives is probably unwise and in need of change. Moreover, it offers evidence for the simplest of all possible alternatives: If you want someone to assess a continuous probability distribution, just ask him to draw it.

Additional work concerned with the effects of elicitation method on the veridicality of subjective judgments was conducted by Stillwell, Seaver, and Edwards [reference 31]. Using an experimental task requiring likelihood ratio estimates, response scales (linear versus logarithmic) and scale endpoints were varied. The results showed that logarithmically spaced scales were superior to the linearly spaced scales. The range of true likelihood ratios was, however, shown to have a strong and significant effect on performance. Subjects were much better able to approximate veridical judgments when less extreme true likelihood ratios were chosen. There was also a significant interaction between endpoint and spacing (logarithmic versus linear) accounting for a relatively large proportion of the variance.

Scale endpoints were shown to influence judgments consistently. Either of two factors may be contributing to this finding. It is possible that the upper endpoint offers an upper bound to responses, thereby limiting the range of values expressed. A second possibility is that the endpoints controlled subjects' judgments about the range in which they could expect the true value to fall. More extreme endpoints may thus produce more extreme responses.

Other research [reference 9] conducted at the University of Southern California during the reporting period focused on isolating the reasons for the well-known finding of conservatism in probabilistic inference tasks (i.e., the tendency to underestimate revised odds given a new piece of evidence). In these experiments, subjects estimated mean log likelihood ratios, already judgmentally aggregated over all data. Such log likelihood ratios processed by means of Bayes' Theorem were found to produce more veridical final odds than posterior odds estimated

directly. In light of general finding of conservatism in probability revision tasks, this suggests that the outputs of formal probabilistic information processing systems are more likely to reflect subjective certainty than are posterior odds judgments. The data also showed that persons using the averaged log likelihood ratio judgments were more orderly in these judgments as evidenced by higher correlation between true final odds and final odds calculated via Bayes' Theorem. It was also found that the diagnosticity of information affected quality of response for both response modes. Estimates became more veridical as the data became more diagnostic. The primary finding of the study was that quality of estimates did not differ significantly in either veridicality or orderliness between likelihood ratio estimates as originally proposed for the PIP technique and the averaged log likelihood method. Both methods were found to produce better estimates than cumulative certainty judgment, as is usual in such comparisons.

These findings also compel a rethinking of the misaggregation explanation of conservatism in probability revision. Mean log likelihood ratio is a judgmentally aggregated response--but it is not conservative (nor yet radical). Apparently, aggregation that has the character of a sum or product (i.e., the target number is outside the range of input quantities) is conservative. Aggregation that has the character of an average (the target number lies near the middle of the range of input quantities) is unbiased. This finding links research on probabilistic inference with a wide variety of other types of research on human judgment.

Research was also conducted during this reporting period on the problem of assessing very low probabilities. This type of probabilistic assessment is of considerable importance since there are many critical low probability/ high expected value decision contexts wherein formal decision

analytic procedures can be beneficially applied. Nuclear engineering has brought this situation to public attention as system failures occur with probabilities typically smaller than 10^{-6} but with values which may exceed 50,000 lives lost. But identical kinds of problems arise frequently in military and political contexts. An obvious example is whether or not a particular limited-war strategy may lead to a widening of the war. Both experimental and applied work have shown, however, that problems arise in the subjective assessment of the likelihood of highly unlikely events.

Research conducted by Lichtenstein, Slovic, Fischhoff, Layman and Combs [reference 22] concerned with the assessment of the probability of lethal events, suggests a remedy for the small probability assessment problem. Instead of direct assessment of the probability of interest, Lichtenstein, et al. asked subjects to judge which of two events was the more likely. They found, with a few notable exceptions, that over 80% of subjects could correctly judge the larger of the probabilities of a pair of events when the ratio of the probabilities was greater than 2:1.

These findings suggest that either a series of comparisons of event pairs or a simultaneous comparison of the event with unknown probability with a list of events with known probability (marker events) may result in significant improvement in probabilistic judgments. Several studies were undertaken to evaluate the potential of this approach. In the first set of these experiments, subjects were asked to place the event of interest into a list of events at a point appropriate to its relative likelihood of occurrence.

Incentive was given to subjects in the form of $3.00 for each response placed in the correct space among 30 spaces between events. Results of this experiment show that subjects were not sufficiently able to perform the task to

warrant its use as an elicitation tool for probabilities. The mean correlation over 120 subjects between response probability (using the midpoint of the response space) and the true probability (using the midpoint of the space in which the event should be placed) was .131. A follow-on study which reduced the subjects' task to a comparison of the event of interest to either one or two other events (as opposed to 31 previously) yielded the same disappointing result. Thus, it does not appear that the marker event method will have applied utility as a method for enhancing the precision of individual estimates of very small probabilities. Detailed analysis of the data reported by Lichtenstein, et al., however, did indicate that the most subjects could correctly identify the most and least likely of two events even though their estimates of absolute and relative probabilities were in error. This suggests that improved accuracy in estimating small probabilities might be achieved by aggregating estimates across individuals.

The above-cited research adds to a growing technical literature on human performance in situations requiring probabilistic judgment. The picture that emerges generally underscores human frailty in this highly important area. With probabilistic judgment a fundamental element of decision behavior in a complex and uncertain world, the research suggests extreme caution in dealing with probability estimates--whether provided by others as a basis for decision, whether formally elicited in decision analytic contexts, or when applied ourselves in our own assessments for decision.

As a companion to the research on biasing factors in probabilistic judgment, the researchers have turned their attention to possible methods of debiasing judgments. Although it was noted that some of the factors that cause poor probabilistic judgments are extremely robust, some

debiasing techniques have been discovered and reported. Kahneman and Tversky [reference 18] prepared a paper during this reporting period which analyzes the judgment process in three tasks that people are often required to perform in forecasting or decision situations: estimating uncertain quantities, assigning probabilities to events, and assessing probability distributions. The Kahneman and Tversky paper considers how these tasks are performed, discusses the major biases that affect each, and suggests procedures that may be useful in eliminating or reducing the biases. Additional work on debiasing was also advanced during the FY 1977 contract period at Perceptronics. Research and analytical efforts directed toward preparation of a major work on methods to improve judgmental accuracy moved ahead as scheduled. The report on this work is to be published during the FY 1978 contract period.

2.1.4 Group assessment of uncertainty - Often, decisions are made not by a single individual but rather several, each of whom should be able to influence the final decision. In decision analysis, a single judgment of uncertainty, as well as a single judgment of value or utility is necessary as input to each branch of the decision-making structure. This apparent incompatiblity has led to a search for techniques which will enable one to derive a valid single expression of uncertainty from multiple judgments. Such research has explored two major strategies, mathematical techniques for the aggregation of individual judgments into a single group estimate, and behavioral techniques which seek group consensus.

Both approaches have mathematical and social psychological difficulties. Dalkey has shown that no formal rule for the aggregation of individual probabilities can satisfy a set of reasonable conditions (such as non-dominance by a single group member). Behavioral techniques likewise

have limitations. Individual group members may concern themselves more with reaching consensus than with the quality of the agreed-on judgment. Individual dominance through personality characteristics or rank within the organization may influence judgments despite its irrelevance to the task.

In an effort to compare various behavioral and mathematical techniques of group probabability assessment Seaver [reference 29] experimentally compared two aggregation rules, weighted arithmetic means and weighted geometric means, and three weighting procedures, equal weights, weights based on self-rating, and DeGroot weights. Five behavioral interaction techniques were compared, the Delphi method, the Nominal Group Technique developed by Delbecq and Van de Ven, a modified nominal group technique in which group members state their estimates and reasons with no discussion, a consensus technique in which groups were to arrive at consensus in any way they wished, and a no interaction or control group in which group members made estimates with no knowledge of other group members' estimates.

The results showed that, in general, interaction among group members reduced differences, reduced the calibration of the judgments, and increased the extremeness of judgments. Therefore, deciding whether or not to use group interaction techniques involves a trade-off between calibration and extremeness of the responses. Although no significant differences were found, minor differences (as well as the results of other studies) point to slight superiority of the nominal group technique to other group interaction methods.

The data show that little if anything is lost by using mathematical techniques rather than behavioral interaction to aggregate individual judgments. Considering the

practical disadvantages of face-to-face meetings of groups, no point exists in bothering with the sometimes lengthy procedures of behavioral interaction. While results of this experiment dealt totally with point estimates, further studies will attempt to elicit continuous distributions.

2.2 Multi-Attribute Utility Assessment Methodology

In the applied portion of the Advanced Decision Technology Program, multi-attribute utility assessment (MAU) methodology has been successfully applied to a variety of important evaluation problems, recent examples being the development of an MAU model for assessing combat readiness of USMC units, and the use of MAU to evaluate alternative system configurations, including alternative mixes of aircraft for naval aviation, combat radio systems, and a number of Army weapon system selection problems. Although MAU methodology has proved useful in its present state of development, there remains considerable room for improvement. In large assessment problems, veridical MAU models can get extremely complex and require elicitation of many values. There is a need to assess the trade-off between model precision and ease of elicitation. There is a need for improved procedures to facilitate value trade-offs in multi-objective decision making. There is also a pressing need for additional work on the validation of MAU procedures. These and other research thrusts pertinent to MAU methodology were pursued during FY 1977.

Leung [reference 20] provided a review of theoretical and empirical research findings regarding the sensitivity of MAU to model specification. The question addressed was whether additional complexities (such as non-additivity, uncertainty, and differential weighting) are useful. Although a few of the studies considered by Leung produced analytic solutions to the questions asked, most were either Monte

Carlo simulations or behavioral studies. The criterion for intermodel agreement, in almost every study, was the correlation between utilities output by the specified models. Leung came to the following conclusions:

1. Additive models should be used as an approximation to more complicated structures, at least for the two-attribute case (unless there are good reasons to believe that a non-additive model is an exact representation of a decision maker's attitudes).

2. Weights do not matter for deterministic additive models when, on the average, the attributes are highly correlated with each other.

3. No conclusions can be drawn regarding how well deterministic models approximated more complicated probabilistic ones.

Of most interest in Leung's analysis was a call for "a measure of robustness other than the coefficient of correlation." Although Leung was referring to studies involving probabilistic models only, the need for additional dependent measures of fit is great.

In a related area, Newman [reference 25] took under study the issue of unit versus differential weighting for additive deterministic utility functions. He exploited the similarity between the formal mathematical structure of the multiple regression model and the additive utility model (under certainty). Using simulation techniques, Newman compared two methods of estimating beta weights for regression models with the unit weighting technique. Both procedures for estimating beta weights, ordinary least squares, (OLS) and ridge regression (RIDGE), proved superior to unit weighting (UNIT) in all cases but one (in this one case, all the true

coefficients were positive, not too far apart, and the sample size was relatively small (50). In the overwhelming majority of cases, unit weighting was simply not appropriate.

Newman also found that the ridge estimates outperformed the OLS estimates a great deal of the time, replicating several studies which demonstrate the superiority of the biased RIDGE procedure to the more popular OLS approach. Newman asserts that the argument in favor of unit weighting is completely shattered when the differential weights are estimated via RIDGE. By replacing the independent and criterion variables in the regression model with the attributes and overall utility construct of an MAU model, one may ask the following question: What subjective estimation procedure do people use in determining their weights for attributes in an MAU? The answer to this question is critical. Unit weighting of attributes in a decision analysis is not appropriate if the decision maker can estimate the attribute weights in a RIDGE or even an OLS manner. However, if subjective estimates of weights are considerably suboptimal, unit weighting is a boon for the application of MAU.

As was pointed out previously, MAU models of complex assessment problems can rapidly become bushy messes requiring the assessment of a great number of values. This circumstance can make application of MAU methodology a time-consuming and costly matter. Further, the voluminous value assessments are subject to error. These factors suggest that there should be a useful trade-off between model complexity (veridicality) and assessment effort and error.

This issue was addressed by Leung [reference 21] in a study designed to explore the possibility of reducing the

number of attributes specified in additive MAU models under certainty.

For each specimen MAU problem used in his experiment, Leung randomly generated a utility array (alternative by attribute) and a set of weights for the attributes. He systematically varied the number of attributes in the full model, the average intercorrelation among the attributes, the number of attributes in the reduced model, and the method for deciding which attributes to eliminate. Leung investigated the following ad hoc procedures for reducing the number of attributes:

1. Retain highest weight attribute, drop the attribute that correlates highest with it; repeat until desired number of attributes are dropped.

2. Ignore intercorrelations, simply drop the desired number of attributes with the lowest weights.

3. Discard lowest weight attribute, retain attribute that correlates highest with it; repeat until desired number of attributes are dropped.

4. Pick highest correlated pair of attributes, discard lower weight attribute among the two, repeat until desired number of attributes are dropped.

Using the distribution (over 1,000 cases) of correlations between the full and reduced model as the dependent measure, Leung found that methods 2 and 3 completely dominated; and he concluded that method 2, considering its ease of application, was the superior procedure. That is, unimportant attributes (attributes which receive small weights) may be eliminated from consideration with little loss. Leung applied this technique to two real-world examples with good results.

The application of the results of this study are subject to the same behavioral questions posed by the Newman study [reference 25] discussed previously. In order to perform Leung's method 2, subjects must be able to accurately rank order the attributes in terms of importance. Whether or not people can do so reliably is a question for further research. Applied work conducted within the program, however, suggests that as a practical matter, rank-ordering performance appears to be good enough.

Additional work on multi-attribute utility assessment was conducted at Stanford University. Research conducted by Keelin [reference 19] focused on the development of fundamental tools for assessing and interpreting preferences over multi-attribute outcomes. Six preference measures were presented as tools for understanding and analyzing multi-attribute preference functions. An easy method for limiting the functional expression of these preferences was developed from a hierarchy of independence conditions and delta properties. The result, summarized concisely in a flowchart, is a simple yet powerful procedure for assessing multi-attribute risk preference for any preference attitude that satisfies deterministic additive independence. Keelin illustrated the procedure by assessing a multi-attribute risk preference function for an actual decision maker.

In related work, Oppenheimer [reference 28] developed an improved method, the proxy approach, to assess a decision maker's preferences over multi-attribute outcomes. Global modeling approaches produce preference functions that are mathematically simple and convenient but often not truly representative of a decision maker's preferences. Local procedures, on the other hand, use sequences of trial solutions to generate linear approximations of preference functions in a slow and inefficient manner. Oppenheimer's research showed that by using global models as proxy functions

in a local procedure, the sequence of trial solutions reaches the optimum faster and without restrictive assumptions. The practical application of this new approach to a planning problem demonstrated the algorithm's success in enabling a decision maker, previously unfamiliar with decision analysis, to assess his multi-attribute preference function.

Additional research conducted at Stanford during FY 1977 focused on the development of an integrated framework for constructing multi-attribute preference functions [reference 23]. Methods for assessing single-attribute preference functions have been well established. The current, state-of-the-art procedure for deriving arbitrary, multi-attribute preference functions, however, has required regularity assumptions to reduce the arbitrariness of the preferences. This technique, called decomposition, has been used because it usually results in a simple and appealing preference model. The difficulty with this approach is seen in the restrictiveness of the two main assumptions:

1. The same preference regularity is imposed on all attributes, using this symmetry to achieve a simple preference function; and

2. Any n-dimensional preference function is decomposed into n one-dimensional preference functions.

In this work, no symmetry assumption is necessary, as each multi-attribute preference function is tailor-fit to only those regularities that exist in a particular problem setting. Those parts of the preference function that are subject to simplifying assumptions are decomposed by using a new classification scheme to derive further independence assumptions for the standard models. Those parts of the preference function that are indecomposable are handled by

using a new discretization scheme along with a behaviorally motivated interpolation rule to fill the gaps. The flexibility of these methods allows an analyst to make trade-offs between the degree of accuracy desired and amount of effort needed.

This integrated framework for decomposable and indecomposable multi-attribute preference functions stands as an important decision analysis aid. The usefulness of the framework was illustrated in an example of the decision to buy a new car. The relationships of the attributes can be assessed in advance, thus allowing an optimal decision to be made in the decision maker's absence.

In related work at Harvard, research focused on the development of interactive algorithms to aid decision makers in making trade-offs in multi-attribute decision tasks. The interactive approach developed requires the decision maker to respond to fairly simple questions in a series of iterations. His responses provide information about his preferences. These trade-off questions are asked in the context of <u>efficient</u> and <u>feasible</u> alternatives only, greatly increasing the realism and relevance for the decision maker. The concept of preferential independence is also applied to make the local assessment of preferences less demanding on the decision maker. In one algorithm, the decision maker is asked to make ordinal paired comparisons (in which only two attributes are varied at a time while the others are held fixed), and his responses are used to evaluate the marginal rates of substitution at the present point. These rates are then used to find a feasible direction of improving preferences.

Yet another simplification of the trade-off-making process is achieved in another algorithm, where at each iteration the decision maker is presented with a two-dimensional

efficient frontier and he has to select the best point on this frontier by a direct analysis of what is feasible and achievable and what is desirable. Work was also begun on the question of how an explicit value function can be constructed during the iterative process. Generally, the idea is to use the decision maker's response to the trade-off questions in order to find the best-fitting additive model and to improve this model as more preference information is gathered.

Detailed algorithms were formulated to solve two types of problems: continuous, in which the set of feasible decisions is continuous and contains an infinite number of decisions, and discrete problems, where the decision variables are treated as discrete (0, 1) variables. A final report on this work is anticipated in January 1978; and working computer programs, evaluated on several representative problems, will be available in June 1978.

Validation of MAU methods has long been a problem meriting attention. While much laboratory work has been done using convergent validation approaches, the more compelling form of validation, comparison of model results to an external criterion, has been elusive. During FY 1977, Eils [reference 8], at the University of Southern California, made an important contribution in research that exploited an external criterion against which to validate multi-attribute utility assessments. Eils elicited utility assessments from 24 groups, each of which consisted of four graduate or upper division undergraduate students who knew each other prior to the experimental session. Group utilities were elicited (via consensus) for ten hypothetical applicants for bank credit cards. The research design completely crossed two factors in assessing group utilities: 1) using a decomposition procedure (MAU) or not, and 2) using a formal group communication strategy (GCS) or not. The quality of each

group's utility judgments was defined to be the Pearson product moment correlation between the group's judged utilities and utilities output from a configural (nonlinear) model used by a major bank in evaluating applicants for Master Charge. A content analysis of the group's verbal interaction was conducted to determine the effects of task structure on the characteristics of the group process. Group satisfaction measures were also obtained.

Eils found that the decision technology of MAU greatly aided groups in reaching decisions that were in some sense consistent with decisions based on a systematic collection and interpretation of a large amount of relevant data (i.e., the bank model). When unit weights were used in place of the elicited differential weights, the MAU groups evidenced even higher correlation with the bank model. The application of a communication strategy did not significantly alter the quality of group evaluations.

Both task interventions (MAU and GCS) significantly influenced the group communication process. In addition, groups employing the MAU did not find the task any more complex or difficult, or any less satisfying than groups not employing the technique. Groups employing GCS did not find their task any less satisfying or complex. Perhaps for the first time in empirical MAU validation research, decomposed judgments have demonstrated a greater degree of fit to an <u>external</u> criterion than holistic judgments. The formalized bank model used to measure judgmental validity reflects the complex nature of the relationship between applicant characteristics and subsequent loan performance evidenced in the data used to generate the formal model. These complex relationships should be similar to the ones inherent in the information which the groups bring to the assessment task in the form of past experience. Thus, the degree to which group decisions correspond to the bank's systematic and

complex evaluation provides a measure of how well the elicitation technique taps the information actually contained in group members' past experience. Eils argues that the MAU procedure he employed proved more valid in the sense that a more complete representation of each individual's past experience was elicited.

Although the criterion for validation used by Eils is not a completely satisfying one, it is certainly better than none at all.

2.3 Research on Bargaining and Negotiation

Many important decisions are concluded in the context of negotiations. During FY 1977, research on this class of decision problem was begun in a program of research planned to span several years. The goal of this research effort is twofold: a) to indicate how formal analytic techniques can be beneficially used in an interactive bargaining situation, and b) to show how formal methods can help in the mediation and arbitration of conflicts.

A review of the vast literature bearing on bargaining and negotiation has begun. The technical analytical literature is, for the most part, esoterically theoretical and not directly applicable to real-world problems. The non-technical, management-oriented literature is largely descriptive and makes little use of formal analytical methods to guide behavior. An important thrust of this research effort will be to bridge the gap between these two general approaches.

Steps have also been taken to create a laboratory setting for direct study of bargaining and negotiation behavior. Students enrolled in a course on Competitive Decision Making are required to play simulated roles in

interactive bargaining situations. This laboratory will provide ample opportunity for experimental intervention in the negotiation process and will provide a substantial statistical record of bargaining behavior. Beyond the laboratory setting, this research effort is also closely coupled to real-world bargaining contexts. Arrangements have been made through Decisions and Designs, Incorporated, to provide access on the part of the Harvard staff to active DoD negotiation problems. In this context, Professor Raiffa has been both an observer and advisor to ISA with regard to the Philippine base rights negotiations which were in progress during part of FY 1977.

Research in this area will be documented in the form of either a major monograph or book rather than as isolated, independent articles. An outline of the planned treatise was completed during the reporting period.

2.4 Crisis Decision Analysis

During FY 1977, significant progress was made toward the objective of developing a normative procedure for modeling crisis decisions. The normative approach prescribes how decisions should be made in contrast to the descriptive approach that explains how they actually are made. The procedure is an efficient algorithm for guiding the modeling techniques of structuring, assessment, and analysis--techniques that parallel the deterministic, probabilistic, and informational phases found in standard decision analysis methods. In a crisis, however, the time available for making decisions is limited, events are often unanticipated and unstructured, and important values are threatened. To help an individual model crisis decisions effectively, standard decision analysis methods can be adapted to:

- Indicate, at any moment, the best decision and the value of further modeling;
- Identify and structure only those variables that may impact the decision;
- Incorporate simple models of important values into early stages of the analysis.

The modeling procedure was tested in a case study of the 1975 Mayaguez ship seizure. The significant accomplishment of this research, as demonstrated in this example, is the successful adaptation of standard decision analysis methods to the unique characteristics of a crisis situation.

2.5 Analytical Efforts

During the past year, eight major analytical papers addressing a variety of issues in decision-aiding technology were completed. The first of these (Fischhoff, reference 11) provides a critical assessment of the use and abuse of a variety of formal decision-aiding methods (cost-benefit analysis, risk assessment, and decision analysis) that are increasingly used in public policy decisions. The paper discusses the rationale for such analyses, their acceptability as guides to decision making, the problems such analyses encounter, the possibilities for misuse, and the steps to be taken to increase their contribution to society.

In this critique, cost-benefit analysis and related techniques are found to have a critical role in guiding decisions, particularly those affecting large segments of our society. Whatever flaws such analyses may have, they are clearly superior to less systematic approaches. It is important, however, for both the analyst and the non-expert consumer of such analyses to understand the errors to which

the analyses are prone in order to maintain a critical perspective. Indeed, the institutionalization of such criticism is advocated. This review also identifies a need for research directed at clarifying psychological (subjective) aspects of the analytic process in order to: a) reduce the errors and omissions that may be made by the analysts and b) improve communication of the results of analyses and the assumptions under which they were reached to decision makers.

In a companion paper, Fischhoff [reference 12] presents a critical analysis of the method and practice of decision analysis. The vehicle for his discussion is an analogy drawn between decision analysis and the somewhat older profession of psychotherapy. Both offer a variety of techniques designed to help people function in a difficult and uncertain environment; both developed rapidly, sustained by a coherent underlying theory and anecdotal evidence of having helped some clients.

The salient conclusions presented by Fischhoff are:

a. Well-developed methodologies exist for evaluating the effectiveness of social interventions (of which both decision analysis and psychotherapy are examples). Combined with pioneering evaluations like those of Brown and Watson at Decisions and Designs, Incorporated, this methodology could give leverage to determining where decision analysis is most cost effective.

b. Substantial progress has been made in assessing the validity of some of the elicitation procedures used by decision analysts. However, relatively little is known about the robustness of these results; i.e., we do not know to what extent they

apply in different contexts and with different decision makers. In addition, the subjective side of some aspects of decision analysis, like how decision problems are structured, has barely been studied.

c. Much of the success of a decision analysis may depend upon the analyst's self-presentation, ability to get along with clients, sensitivity to clients' unstated desires and uncertainties, and capacity for instilling confidence. A manual of advice on how to fulfill these functions would be useful.

d. Highly competent analyses can fail as guides to decisions if they adopt too narrow a definition of the decision problem. One must consider the political, organizational and legal constraints which can make a technically feasible course of action socially unfeasible. Ways are needed to incorporate into analyses the possibility that selected courses of action will not be adopted at all, or at least not as planned.

e. Decision analysts might usefully consider following the example of psychotherapists in developing some sort of association that would monitor how analysts are trained and how analyses are performed, in order to protect the profession, clients, and the public from substandard work.

Another analytical paper (Brown and Watson, reference 4), published during the past 12 months, presents a discussion of alternative methods for testing management systems (decision systems) in a defense context. The authors point out that

the general principles of sampling underlie all approaches to management system testing; they present a discussion of the pros and cons attendant to different testing approaches ranging from classical experimentation, simulation and prototype testing through clinical observation and intuitive assessment. Although there are deficiencies with all the methods discussed, Brown and Watson conclude that the prototype testing method, as exemplified in the engineering design process, is the most promising method for testing management systems.

In another area, O'Connor and Edwards [reference 26] present a detailed discussion of the use of scenarios in the evaluation of complex systems. They point out that the use of carefully selected scenarios (specified futures) is essential in evaluating complex systems since the exhaustive projection of all possible outcomes of system deployment poses an intractible problem. They note that scenarios must be selected to satisfy two objectives: they should validly represent the future world in which the systems under consideration will be deployed, and they should discriminate between the alternative systems under consideration in terms of the utility of their respective deployment. The paper focuses on the definition and characterization of the scenario problem, the proper selection of scenarios, the evaluation of scenario probabilities, and the use of scenarios in system design and in choosing among a specified set of alternative systems. This paper is of particular significance in that it is the first critical assessment of the vital role of scenarios as a basis for decision and should constitute a valuable point of departure for further research.

With the foregoing scenario work as a partial point of departure, O'Connor [reference 27] presents a lengthy discussion of procedures for assessing the value of command and control (C^2) capabilities. The paper details the application

of multi-attribute utility assessment methodology as a means of arriving at a relative assessment of the worth of alternative C^2 concepts or systems.

In a related analytical paper, Gardiner and Ford (1977) explain a technique for evaluating the results of computer simulation models of social systems with additive, riskless, multi-attribute utility functions. Computer simulation models are frequently developed and used as policy analysis tools that show, for the system being modeled, how its behavior over time is influenced, if at all, by proposed policies. Many simulation efforts stop at this point and leave the synthesis of the derived results to unaided intuitive approaches. The emphasis and focus is on developing models that show consequences of policies, not on formally evaluating these consequences. Consequently, simulation models and accompanying policy recommendations are frequently criticized for failing to take into account societal interests and values. This paper discusses how MAU can be applied to the output of computer simulations to remedy the deficiencies inherent in the system dynamics methodology.

Elaboration is made to an application in energy boom towns where a system dynamics model of a boom town "feeds" evaluation models developed for nine viewpoints of individuals in Framington, New Mexico, including that of the mayor, a conservationist, representatives of the energy industry, etc. The applicability of this merged technique to military boom-town phenomena is discussed as well as its application to military "bust towns", i.e., those instances where U.S. military installations are closed.

A technical report prepared by Edwards [reference 7] presents an interesting and valuable discussion of the problems with and prospects for institutionalizing decision-analytic techniques in Federal bureaucratic contexts, and is

in particular responsive to the views on that topic of Mr. Joseph Coates, of the U.S. Office of Technology Assessment. As examples, Edwards summarizes two extensive applications, both using ARPA-developed technology but neither funded by ARPA. Both studies include technical innovations highly relevant to ARPA and DoD needs and problems. The technical report makes special effort to be readily understandable by those unacquainted with decision technology, probability, and the like; it does, however, assume experience with Federal bureaucracies.

The first example outlined by Edwards involves the technology of probability assessment and use of Bayes' Theorem. Probabilities of various diagnoses were assessed by clinicians in emergency room settings all over the U.S. to determine the diagnostic efficacy of the radiographic procedures employed. Specifically, clinicians provided probability estimates of the most likely diagnosis and most important diagnoses before and after interpretation of about 8,000 x-rays. The log likelihood ratio, computed from the prior and posterior probabilities assessed, served as a measure of the influence of x-rays on clinical diagnosis. The assessments were accomplished "in the field" by clinicians with a minimum of technical training.

The example suggests extensions of the described methodology to a variety of real-world settings. Any situation in which a costly, perhaps dangerous, procedure to gather information is employed is amenable to this investigatory approach. As technology in general advances and methods to reduce uncertainty become increasingly more available, the decision of whether the amount of additional information obtained is worth the energy expended in gathering it will become both more important and complex. As technological sophistication increases, the stakes increase; and the intuitive ability of man to choose beneficially between

seeking or not seeking more information decreases. Thus, techniques of studying the efficacy (in some sense) of information collecting procedures (such as radiographs) will become increasingly important. The most obvious military example has to do with collection of intelligence information.

In the second example, the technique of multi-attribute utility analysis (MAU) is applied to a highly complex social decision-making problem: siting a nuclear waste disposal facility. In contrast to the first example, the primary focus is on determining measures of value, not uncertainty. The most important feature of this application is the use of the MAU procedure (developed for use by individuals) by a face-to-face group of decision makers.

Group interactions were structured around the MAU tasks of determining dimensions of importance, and weighting those dimensions. Experts in nuclear engineering from several countries comprised the groups. Hypothetical alternative waste disposal sites were generated by one of the experts who had extensive experience with the siting problem. A numerical demonstration of the MAU evaluation of sites was performed, using the expert-assessed weights and linear transformations of values (or log values) as location measures on utility curves.

Two additions to the usual MAU technique were employed. Rather than obtaining ratio scaled weights in which only ratios involving the least important attribute are checked, the respondents were required to judge ratios of all possible pairs of weights. This change in elicitation procedure probably enhanced the reliability, and hence the validity, of the utility model parameter estimates determined via the weighting procedure.

A final analytic paper completed during FY 1977 was a review of objective and subjective evidence bearing on the efficacy of decision-aiding concepts and decision aids [reference 10]. The review indicates that while there is a great deal of piecemeal laboratory research that supports the principles underlying decision-aiding concepts, there is little objective evidence from real-world applications as to the genuine utility of decision aids. Anecdotal evidence, however, indicates that users of decision aids have apparently found them useful. It is a healthy posture, though, to view anecdotal results with a degree of skepticism. The lengthy draft report reflecting the reviewed literature on this topic is presently undergoing editorial review in preparation for printing.

2.6 Pilot Applications

Seven pilot applications were carried out during FY 1977 as a means of evaluating applied research, generating ideas for new research, and as a step in transferring research products to end users. The most significant of these applied efforts was a program carried out (and still ongoing) with the operations (J-3) and intelligence (J-2) staffs at Headquarters, U.S. European Command (HQEUCOM), and at the subordinate European Commands. This project, arising out of a series of briefings given at Headquarters, EUCOM, during FY 1976, was significant both because the types of decision problems addressed were substantively different from those encountered previously in other applications work and because it represented the first time that work had been carried out in the field with a major command.

2.6.1 EUCOM - The EUCOM project was carried out by transferring an IBM 5100 to Headquarters, EUCOM, and to each of its subordinate Commands. Members of the ARPA project team visited those headquarters at approximately six-week

intervals to instruct personnel in the use of decision-analytic software and to assist them in applying the software to problems of current concern to the Commands.

Two applications software packages were developed with the J-2, one for constructing and evaluating influence diagrams and one for constructing Bayesian hierarchical models; and two for the J-3, one for rapid option screening in the face of uncertainty and the other a general multi-attribute utility package. In all, EUCOM and subordinate Command personnel used the software to construct some 45 models over a 12-month period. Approximately one-half of these models produced results which were incorporated into EUCOM recommendations or studies. Two of the models, one for the J-3 and one for the J-2, were of major significance and were extensively briefed both in-theatre and in the Washington area. As a consequence of the substantive contribution which the software made to the resolution of European Command problems, the work, which was originally scheduled to terminate in April 1977, was extended through March of 1978.

A number of conclusions can be developed from the EUCOM experience. Paramount among these was the demonstration that advanced decision-analytic techniques could be transferred to operational commands in spite of the substantial constraints imposed by the operational environment. Evidence that transfer occurred is provided by the fact that personnel in the using Commands carried out a substantial number of analyses on their own, and by a number of observations that plans and studies by the Commands were beginning to reflect a decision-theoretic structure. In this regard, it was apparent that the IBM 5100 and the applications software were important ingredients in the transfer process. The software provided both an elicitation aid and a reminder to users of the key elements required by decision theoretic

structure, that is, the structuring of a problem into acts, events, criteria, probabilities, and utilities. In this regard, one user remarked that as a result of intensive use of the applications software early in the program, he now found the act of structuring a problem was often sufficient to obtain a solution without a formal analysis using the computer.

At the beginning of the EUCOM project, the deputy CINCEUR stated that one of his most important problems was that of achieving battlefield integration, that is, one of achieving relevant communications across the elements of the joint staff. In retrospect, the applications software made a significant contribution to this process. For example, the rapid-option-screening package both provided a place for, and required separate inputs by, the J-3 and the J-2. In addition, it allowed members of the staff working on the same problem to establish very clearly what the options were, what the criteria were, what the critical uncertainties were; and it required as part of the structuring process that members of the problem team distinguish among fact, assumption, and judgment. The importance of these apparently simple distinctions was underscored time and again when, as a result of making them, new insights were obtained. Another problem, similar in nature, was determining how to integrate the inputs from the component J-2's, such that maximum use was made of their air, ground, and naval specialization. The Bayesian hierarchical model turned out to be a very logical and compelling solution to this problem.

As an ancillary benefit, the J-3 and the J-2 found that more effective briefings could be given by organ izing them around the computer-based decision aids rather than by using the more traditional briefing techniques. In effect the computer could be used to display frame by frame

many of the same items that could also be displayed with vugraphs, but with the additional benefit that, unlike the more static presentation, it allowed those in the audience to modify certain assumptions and to determine the implications for choice. This capability addresses a deficiency in the normal briefing procedure which has been endemic all along.

With respect to technology transfer, the kinds of products which appealed to the J-3 in general did not appeal to the J-2, although the reverse was not necessarily the case. The J-3 personnel seemed to be generalists and were more interested in possessing a <u>model building</u> capability than in working with models which had been constructed by someone else and were simply available for them to use. The J-2, on the other hand, who tend to be more substantively specialized, were less interested in knowing how to build models but were highly desirous of models which had been built and which were specifically tailored to their problems.

A number of research issues were identified which can provide the basis for both additional case studies and basic research. In the applied area, a number of software modifications are required to facilitate user interaction. In particular, the software needs to be modified so that it provides more of a tutorial function than is currently the case. The use of the software for briefing highlighted the need to capture in qualitative form the rationale for many of the judgments that went into a model. Because the inputs often were obtained from a variety of people, it was very difficult for any one person to be sufficiently familiar with the model to readily explain in a briefing the reasons for each input to a model. Accordingly, the software has been modified to provide a rationale-capturing capability.

As has been mentioned before, there is broad scope for research on the problems of option generation, problem structuring, and probability and utility elicitation. In particular, the process of option generation was observed to occupy in many situations by far the majority of the analysts' time, leaving very little time for option evaluation. For this reason, it was often the case that analysts seemed to use a satisficing principle whereby if an option were generated that satisfied some minimal set of constraints, it would be acted upon, and no additional options would be generated and evaluated. Finally, it is necessary to carry out research to develop self-contained instructional material so that the Commands develop an organic capability for training.

Although it is difficult to accumulate definitive, objective data on the effectiveness of the decision-aiding technology that served as the vehicle for EUCOM pilot application activities, our sense of the situation leads us to believe that user reaction to the aiding technology has been generally favorable. Perhaps the most persuasive evidence of favorable response to the aiding concepts is the fact that Headquarters, EUCOM has issued a formal <u>Required Operational Capability (ROC)</u> statement detailing a need for decision aids and calling for continued research and development support toward that end.

2.6.2 <u>Multi-attribute utility applications</u> - A limited number of pilot applications were carried out during previous years to apply multi-attribute utility techniques to problems of evaluating design alternatives for system acquisition in a design-to-price environment. This year, those prior seeding efforts achieved fruition. Multi-attribute utility evaluation techniques developed as a consequence of the initial pilot applications are now being used on two Navy

procurements and three Army procurements. In addition, the technology has been briefed to all Army program managers of major system acquisition programs as the technique which they should consider using as a means of implementing the new Department of Defense procurement directives. The methodology has also been proposed by the WWMCCS system engineer as the methodology which will be used in a multi-year, multi-million-dollar study to evaluate the worth of ADP in the WWMCCS system. As was the case in EUCOM, much of the success of this technology transfer can be attributed to the user-oriented application software implemented on the IBM 5100 portable computer.

During FY1977, a number of pilot applications were undertaken using the basic MAU technology developed for systems acquisition. However, these new applications were in areas substantively quite different from systems acquisition. The first involved the development of a model that can be used by the Air Force Tactical Air Command to develop force requirements in a way that is responsive to the mission-oriented planning directives now being implemented throughout the Department of Defense. In this pilot application, the hierarchical multi-attribute utility evaluation concept was combined with a management information system to allow various levels of command to make trade-offs appropriate to their particular level of command and to retrieve on-call the rationale underlying judgments which had been incorporated in the model by subordinate levels. As it was implemented, the technique used the numbers as a means of summarizing information and also as an indexing system to allow retrieval of specific qualitative statements containing rationale which most effectively discriminated among the alternatives. In this way, the information reduction which often takes place in an organization is accomplished by using numerical summaries, and the rationale underlying the data in the model is preserved intact. The prototype of this combined

evaluation model and management information system has been briefed both to the Commanding General of the Tactical Air Command and to the Commander, Air Force Systems Command. It was also described in a paper presented at the National Defense University and in a paper presented at the Military Operations Research Society's spring meeting. The Air Force currently plans to implement the technique on a pilot basis in FY 1978.

A second multi-attribute utility model extension was an intensive short-term effort involving direct interaction with Marine Corps personnel to develop a model for use by the Marine Corps in assessing the combat readiness of Marine Corps units. Based on a structure and a set of attributes defined in preliminary fashion by the Marine Corps, and encompassing over 800 criteria bearing on combat readiness, the model permits a more comprehensive, valid and rapid evaluation of readiness than any method heretofore employed. The model was applied on a trial basis within the Marine Corps, and highly favorable user reaction led to its formal adoption by the Marine Corps. Planning is now underway to implement the model on a time-sharing computer system at Marine Corps headquarters, and subsequent work in FY 1978 will be directed toward extending the model to include additional criteria which are more difficult to quantify.

The third MAU application was to develop a decision aid that could be used by Marine Corps personnel to make critical resource allocation choices as reflected in the Program Objectives Memorandum (POM) process. These choices are of critical importance in achieving maximum military effectiveness within a finite budget. The resultant decision aid was used extensively by Headquarters personnel in preparing the USMC POM for fiscal year 1979.

The POM decision aid enabled the responsible staff element to present well structured decision options to the Commandant for his consideration. The structured decision options made clear the trade-offs among equipment, personnel, and life-cycle costs, and the implications of these to current and future capabilities. Beyond its on-line utility in screening possible elements for the POM, the model yielded an additional gain by highlighting for high-level Marine Corps decision makers the difference in overall utility that results in basing allocation decisions on consideration of units of maximum benefit within resources as opposed to making the allocation on the basis of units of cost/benefit. The former approach, which had been the accepted practice, yields suboptimal allocation choices relative to the cost/benefit criterion.

The POM model has been briefed extensively within Headquarters, Marine Corps and is now receiving considerable attention from a number of Navy personnel. During FY 1978, it is planned to extend the current methodology to assist the Marine Corps in zero-base budgeting for the development of the 1980 POM. At the end of FY 1978, the methodology will be transferred to Headquarters, Marine Corps computers. A technical report [reference 6] outlining the rationale underlying the model has been completed and is now in press.

These new applications, together with those completed in previous years, highlight the robust quality of multi-attribute utility technology. A common characteristic, however, shared by all of the examples described above, and retrospectively, by examples completed earlier, is the need to provide data which supports the quantitative assessments of the multi-attribute utility models. Other forms of decision-analytic modeling normally make use of relatively few inputs compared with those required by the MAU technology.

Thus, it is not unreasonable that a single individual can become sufficiently knowledgeable so that he can describe the reasons underlying the various assessments in the smaller models. The larger MAU models, however, requiring as they do many hundreds of assessments and involving numerous people who are experts in different areas, are beyond the scope of any one individual to grasp. Therefore, in developing qualitative arguments which can support the conclusions summarized by the output of the model, it is necessary that users have available in the computer the rationale imputed by all the different experts when making their assessments. To a limited extent, this technique has been evaluated in the context of the pilot application work conducted over the past year; and in each instance, it has not only enhanced the utility of the technology, but it has substantially improved its credibility in the eyes of those not familiar with it on a day-to-day basis.

Future work in this area, we believe, should address the problem of refining rationale-capturing and should focus on extending the scope of applications to other parts of the procurement process to capture, at one end, the bargaining process which accompanies the development of requirements of a new weapon system and, at the other extreme, development of techniques for long-term DoD capital investment decisions. The latter problem is currently receiving research attention at both Harvard and Stanford under the auspices of the Advanced Decision Technology Program.

Negotiations. The pareto-optimal negotiation model, which was pilot-tested approximately two years ago, showed great potential as an aid to negotiators, but at that time was impractical for them to use because of the large computer required to implement the optimization algorithm. In FY 1977, a simplified algorithm which calculates a linear approximation to the pareto-optimum surface was developed.

This new innovation permitted the previously complex model to be implemented on an IBM 5100 portable computer for on-line application. As a consequence, negotiators attached to OSD(ISA) used the technique on a continuing basis over a three-month period on an active and important Department of Defense negotiating problem. The participants stressed the utility of the negotiating model not only as an aid in evaluating possible treaty options, but as a facilitator to obtain agreement on different positions among the various constituencies represented in the U.S. negotiating team.

While a substantial step forward has been taken in providing an aid for negotiators, it is quite clear that a great deal of additional work remains to be done. For example, the current negotiating aid can be used to assess the utility of various treaty packages and to define a pareto-optimal set of treaties, but it provides no assistance in guiding the negotiating process itself; that is, it provides no way to identify which issues should be addressed in the initial stages of the negotiation and which ones should be deferred for resolution at a later time. To make progress on these research issues, an informal working group has been established consisting of both theoreticians and practicing negotiators. A member of this group is W. J. Usery, former Secretary of Labor and a mediator of international repute. Quarterly meetings are held with this group to review progress to date on practical issues and to obtain insight into the characteristics of the negotiating process which could be used to guide further theoretical developments. In addition, a major research effort is underway at Harvard concerned with problems of decision making via bargaining and negotiations.

National Defense University. A number of lectures, seminars, and mini-analyses were undertaken during FY 1977 with faculty and students of the National Defense

University (the National War College, and the Industrial College of the Armed Forces). The intent of the mini-analyses was both to demonstrate to the students and to the faculty a new problem-solving methodology that could be incorporated into the NDU curriculum and to stimulate interest in the technology on the part of the students so that they could apply it to problems when they return to their parent organizations. We believe that these objectives were achieved. The students participated in the mini-analyses with great enthusiasm and the results were briefed to high-level Department of Defense decision makers. The impact of pilot application activities at the National Defense University is evidenced, in part, by follow-up inquiries received from students who, having left the National Defense University, are now interested in trying out the methodology on problems in their parent organizations.

It is apparent that the transfer of advanced decision technology must rely heavily on case study applications. However, it is also clear that much can be done to introduce the technology in an academic environment by using these real-world case studies as a teaching vehicle. Curricula need to be developed not only for the National Defense University but for the Armed Forces Staff Colleges and Service academies so that students are exposed to the fundamental principles of decision analysis at at least three stages in their careers. Plans have been made to pursue such curriculum development under the FY 1978 program.

As a matter related to technology transfer and curriculum development, it is worth noting that the draft <u>Handbook for Decision Analysis</u> [reference 1], prepared under ARPA and ONR support, was extensively revised and refined into final form during FY 1977 and is now in press. Copies should be available for distribution in December 1977. The handbook is a clear exposition of the principles and appli-

cation of decision analysis methodology. Wide distribution is planned, and we expect that such dissemination will play a major, highly leveraged role in gaining understanding and acceptance of advanced decision-aiding concepts.

DIA. In the early days of the advanced decision technology program, the bulk of the applications work was centered on case studies with analysts in the Defense Intelligence Agency. After some five years of work, it is now apparent that these initial efforts are paying large dividends. During FY 1977, the Defense Intelligence Agency committed itself to probabilistic estimation, and a major training program is now underway to train several hundred analysts in probability encoding. In addition, a small group of ten analysts is working with an IBM 5100 and some probability model-building software as an initial step in incorporating probabilistic modeling into DIA. They are also working with the Bayesian hierarchical software to develop advanced tools for indications and warning analysis.

Higher Order Language Assessments. A final pilot application activity conducted during FY 1977 was concerned with the assessment of the worth to DoD of shifting to the use of a single higher order language (HOL) for use in the development or support of application software for "embedded" computers in DoD systems. An embedded computer is one which is an integral part of a larger special-purpose system (e.g., a fire control system, guidance system) as opposed to general-purpose computers used, for example, for accounting or scientific applications.

Within DoD today, there are over 450 HOL's in use. The proliferation of HOL's has resulted in a host of inefficiencies in software development and maintenance. In light of the apparent costs generated by the multitude of HOL's, ARPA initiated an effort to determine the degree to

which reducing the number of such languages would yield cost savings. There appeared to be overwhelming intuitive evidence to support the conclusion that adopting a single HOL for all future embedded computer applications would be the best course of action. However, attempts to support that position with quantitative arguments proved impossible because of a paucity of data. In light of this, ARPA supported a DDI effort to apply a structured, subjective decision analytic approach to the problem.

Using or adapting existing computer-based decision-aiding models, DDI undertook an analysis of the problem. Fourteen different criteria were identified for evaluating contending HOL's, and a model was developed which would permit assessment of different strategies (mixes of languages, date of introduction, acceptance rate, cost savings, etc.) for shifting to alternate HOL postures. These models were made available to the DoD HOL Steering Group for their use in resolving the HOL matter, and we understand that they have been well received.

REFERENCES

1. Barclay, S., Brown, R. V., Kelly, C. W., Peterson, C. R., Phillips, L. D., and Selvidge, J. Handbook for Decision Analysis. Technical Report TR 77-6-30. McLean, Virginia: Decisions and Designs, Incorporated, September 1977.

2. Bar-Hillel, M. The base-rate fallacy in probability judgments. Technical Report PTR-1042-77-4/1. Eugene, Oregon: Decision Research, a Branch of Perceptronics, April 1977.

3. Beyth-Marom, R. and Fischhoff, B. Direct measures of availability and judgments of category frequency. Technical Report PTR-1042-77-3. Eugene, Oregon: Decision Research, a Branch of Perceptronics, March 1977.

4. Brown, R. V. and Watson, S. R. Testing procedures in the design of a management system: Some methodological reflections. Technical Report TR 77-1. McLean, Virginia: Decisions and Designs, Incorporated, February 1977.

5. Buede, D. M. Assessment of a joint probability distribution on an evolving dynamic process. Technical Report EES DA 77-5. Stanford, California: Stanford University, September 1977.

6. Buede, D. M. and Peterson, C. R. An application of cost-benefit analysis to the USMC Program Objectives Memorandum. Technical Report TR 77-8-72. McLean, Virginia: Decisions and Designs, Incorporated, November 1977.

7. Edwards, W. Technology for Director Dubious: Evaluation and decision in public contexts. SSRI Research Report 77-2. Los Angeles: Social Science Research Institute, University of Southern California, August 1977.

8. Eils, L. C. The effects of communication strategy and decision technology on the process and product of decision-making groups. Doctoral Dissertation, University of Southern California, June 1977.

9. Eils, L. C., Seaver, D. A., and Edwards, W. Developing the technology of probabilistic inference: aggregating by averaging reduces conservatism. SSRI Research Report 77-3. Los Angeles: Social Science Research Institute, University of Southern California, August 1977.

10. Fischer, G., Edwards, W., and Kelly, C. W. *Decision* theoretic aids for inference, evaluation, and decision making: A review of research and experience. McLean, Virginia: Decisions and Designs, Incorporated, in press.

11. Fischhoff, B. The art of cost-benefit analysis. Technical Report PTR-1042-77-2. Eugene, Oregon: Decision Research, a Branch of Perceptronics, February 1977.

12. Fischhoff, B. Decision analysis: Clinical art or clinical science? Technical Report PTR-1042-77-5. Eugene, Oregon: Decision Research, a Branch of Perceptronics, May 1977.

13. Kahneman, D. and Tversky, A. Prospect theory: An analysis of decision making under risk. Technical Report PTR 1042-77-4. Eugene, Oregon: Decision Research, a Branch of Perceptronics, April 1977.

14. Fischhoff, B., Slovic, P., and Lichtenstein, S. Fault trees: Sensitivity of estimated failure probabilities to problem representation. Journal of Experimental Psychology. In press.

15. Gardiner, P. C. and Ford, A. A merger of simulation and evaluation for applied policy research in social systems. In Snapper, K. (Ed.), Models in Metrics for Decision Makers. Washington, D.C.: Information Resources Press, in press.

16. Grum, A. F. Information perishing and replenishment. Technical Report EES DA 77-1. Stanford, California: Stanford University, October 1977.

17. John, R. S. and Edwards, W. Subjective probability assessment and the shape of the distribution. Los Angeles: Social Science Research Institute, University of Southern California, in press.

18. Kahneman, D. and Tversky, A. Intuitive predictions: biases and corrective procedures. Management Science, in press.

19. Keelin, T. W. A protocol and procedure for assessing multi-attribute preference functions. Technical Report EES DA 77-2. Stanford, California: Stanford University, July 1977.

20. Leung, P. Sensitivity analysis of the effect of variation in the form and parameters of a multi-attribute utilty function: A survey. Working paper. Los Angeles: Social Science Research Institute, University of Southern California.

21. Leung, P. The effects of reducing the number of attributes in additive multi-attribute utility modeling under uncertainty. Working paper. Los Angeles: Social Science Research Institute, University of Southern California.

22. Lichtenstein, S., Slovic, P., Fischhoff, B., Layman, M., and Combs, B. Perceived frequency of lethal events. Technical Report 76-2. Eugene, Oregon: Decision Research, a Branch of Perceptronics, December 1977.

23. Nahas, K. H. Preference modeling of utility surfaces. Technical Report EES DA 77-4. Stanford, California: Stanford University, December 1977.

24. Nafeh, J. Markov decision processes with policy constraints. Technical Report EES DA 76-3. Stanford, California: Stanford University, April 1976.

25. Newman, R. J. Differential weighting for prediction and decision making studies: A study of ridge regression. SSRI Research Report 77-1. Los Angeles: Social Science Research Institute, University of Southern California, August 1977.

26. O'Connor, M. F. and Edwards, W. On using scenarios in the evaluation of complex alternatives. Technical Report DT/TR 76-17. McLean, Virginia: Decisions and Designs, Incorporated, December 1976.

27. O'Connor, M. F. Procedures for assessing the value of command and control capabilities. Technical Report TR 77-4. McLean, Virginia: Decisions and Designs, Incorporated, July 1977.

28. Oppenheimer, K. R. A proxy approach to multi-attribute decision making. Technical Report EES DA-77-3. Stanford, California: Stanford University, October 1977.

29. Seaver, D. A. How groups can assess uncertainty: Human interaction versus mathematical models. Los Angeles: Social Science Research Institute, University of Southern California, in press.

30. Shaklee, H. and Fischhoff, B. Limited minds and multiple causes: Discounting in multicausal attributions. Technical Report 77-11. Eugene, Oregon: Decision Research, a Branch of Perceptronics, in press.

31. Stillwell, W. G., Seaver, D. A., and Edwards, W. The effects of response scales on likelihood ratio judgments. SSRI Research Report 77-5. Los Angeles: Social Science Research Institute, University of Southern California, August 1977.